

11-1-2022

Monochromatic Globular Clusters as a Critical Test of Formation Models for the Dark Matter-deficient Galaxies NGC 1052-DF2 and NGC 1052-DF4

Pieter van Dokkum
Yale University

Zili Shen
Yale University

Aaron J. Romanowsky
San Jose State University, aaron.romanowsky@sjsu.edu

Roberto Abraham
University of Toronto

Charlie Conroy
Harvard-Smithsonian Center for Astrophysics

See next page for additional authors

Follow this and additional works at: https://scholarworks.sjsu.edu/faculty_rsca

Recommended Citation

Pieter van Dokkum, Zili Shen, Aaron J. Romanowsky, Roberto Abraham, Charlie Conroy, Shany Danieli, Dhruva Dutta Chowdhury, Michael A. Keim, J. M. Diederik Kruijssen, Joel Leja, and Sebastian Trujillo-Gomez. "Monochromatic Globular Clusters as a Critical Test of Formation Models for the Dark Matter-deficient Galaxies NGC 1052-DF2 and NGC 1052-DF4" *Astrophysical Journal Letters* (2022).
<https://doi.org/10.3847/2041-8213/ac94d6>

This Article is brought to you for free and open access by SJSU ScholarWorks. It has been accepted for inclusion in Faculty Research, Scholarly, and Creative Activity by an authorized administrator of SJSU ScholarWorks. For more information, please contact scholarworks@sjsu.edu.

Authors

Pieter van Dokkum, Zili Shen, Aaron J. Romanowsky, Roberto Abraham, Charlie Conroy, Shany Danieli, Dhruva Dutta Chowdhury, Michael A. Keim, J. M. Diederik Kruijssen, Joel Leja, and Sebastian Trujillo-Gomez



Monochromatic Globular Clusters as a Critical Test of Formation Models for the Dark Matter-deficient Galaxies NGC 1052-DF2 and NGC 1052-DF4

Pieter van Dokkum¹ , Zili Shen¹ , Aaron J. Romanowsky^{2,3} , Roberto Abraham⁴ , Charlie Conroy⁵ , Shany Danieli^{6,9} , Dhruva Dutta Chowdhury¹ , Michael A. Keim¹ , J. M. Diederik Kruijssen⁷ , Joel Leja⁸ , and Sebastian Trujillo-Gomez⁷

¹ Astronomy Department, Yale University, 52 Hillhouse Avenue, New Haven, CT 06511, USA

² Department of Physics and Astronomy, San José State University, San Jose, CA 95192, USA

³ Department of Astronomy and Astrophysics, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95064, USA

⁴ Department of Astronomy & Astrophysics, University of Toronto, 50 St. George Street, Toronto, ON M5S 3H4, Canada

⁵ Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA, USA

⁶ Department of Astrophysical Sciences, 4 Ivy Lane, Princeton University, Princeton, NJ 08544, USA

⁷ Astronomisches Rechen-Institut, Zentrum für Astronomie der Universität Heidelberg, Mönchhofstraße 12–14, D-69120 Heidelberg, Germany

⁸ Department of Astronomy & Astrophysics, Penn State University, 525 Davey Laboratory, University Park, PA 16802, USA

Received 2022 July 14; revised 2022 September 4; accepted 2022 September 16; published 2022 November 15

Abstract

It was recently proposed that the dark matter-deficient ultradiffuse galaxies DF2 and DF4 in the NGC 1052 group could be the products of a “bullet dwarf” collision between two gas-rich progenitor galaxies. In this model, DF2 and DF4 formed at the same time in the immediate aftermath of the collision, and a strong prediction is that their globular clusters should have nearly identical stellar populations. Here we test this prediction by measuring accurate $V_{606} - I_{814}$ colors from deep HST/ACS imaging. We find that the clusters are extremely homogeneous. The mean color difference between the globular clusters in DF2 and DF4 is $\Delta_{\text{DF2-DF4}} = -0.003 \pm 0.005$ mag, and the observed scatter for the combined sample of 18 clusters with $M_{606} < -8.6$ in both galaxies is $\sigma_{\text{obs}} = 0.015 \pm 0.002$ mag. After accounting for observational uncertainties and stochastic cluster-to-cluster variation in the number of red giants, the remaining scatter is $\sigma_{\text{intr}} = 0.008^{+0.005}_{-0.006}$ mag. Both the color difference and the scatter are an order of magnitude smaller than in other dwarf galaxies, and we infer that the bullet scenario passes an important test that could have falsified it. No other formation models have predicted this extreme uniformity of the globular clusters in the two galaxies. We find that the galaxies themselves are slightly redder than the clusters, consistent with a previously measured metallicity difference. Numerical simulations have shown that such differences are expected in the bullet scenario, as the galaxies continued to self-enrich after the formation of the globular clusters.

Unified Astronomy Thesaurus concepts: [Dark matter \(353\)](#); [Dwarf galaxies \(416\)](#); [Globular star clusters \(656\)](#)

1. Introduction

NGC 1052-DF2 (or DF2) and NGC 1052-DF4 (DF4) share two unusual properties that set them apart from all other known galaxies. First, their globular clusters are, on average, a factor of ≈ 4 brighter and a factor of ≈ 2 larger than canonical values (van Dokkum et al. 2018a, 2019; Ma et al. 2020; Shen et al. 2021a). Furthermore, their velocity dispersions are consistent with their stellar mass alone and much smaller than expected from a normal dark matter halo (van Dokkum et al. 2018b; Wasserman et al. 2018; Danieli et al. 2019; Emsellem et al. 2019).

Initially, the main question was whether the dynamical masses and distances were measured correctly (see, e.g., Martin et al. 2018; Laporte et al. 2019; Trujillo et al. 2019), but as the anomalous properties of the galaxies were gradually confirmed (and corroborated with independent evidence; see Dutta Chowdhury et al. 2019; Keim et al. 2022), the focus shifted to the question of how they were formed. Proposals include assembly out of tidally removed gas (Fensch et al. 2019),

stripping of dark matter by close encounters with NGC 1052 (Ogiya 2018; Carleton et al. 2019; Nusser 2020; Moreno et al. 2022; Ogiya et al. 2022) or NGC 1035 (Montes et al. 2020), jet- or outflow-induced star formation, like Minkowski’s object (van Breugel et al. 1985; Natarajan et al. 1998), and extreme feedback in low-mass halos (Trujillo-Gomez et al. 2022).

Silk (2019) suggested that DF2 could be the result of a “mini bullet cluster” event, where the dark matter and the baryons became separated in a nearly head-on encounter between two gas-rich progenitor galaxies. Such a collision produces two dark matter-dominated remnants with a globular cluster-rich, dark matter-free object in between them that formed from the shocked gas. This scenario was further explored with simulations in Shin et al. (2020) and Lee et al. (2021), who showed that collisions between an unbound object and a satellite could explain many of the observed properties of the galaxies and occur with some regularity in cosmological simulations. The main issue with this model—as with many of the alternative explanations listed above—is the presence of two dark matter-deficient galaxies, seemingly requiring light- to strike twice in the same group.

Recently, we suggested that a single bullet dwarf collision may have produced both DF2 and DF4 (van Dokkum et al. 2022). This joint formation is consistent with the striking similarities between the two galaxies, their radial velocities and line-of-sight distances, the emergence of multiple clumps in at

⁹ NASA Hubble Fellow.



least some bullet collision simulations (Shin et al. 2020), and with the discovery that DF2 and DF4 are part of a remarkable trail of ≈ 10 galaxies in the group.

As noted in van Dokkum et al. (2022), this formation model for DF2 and DF4 is falsifiable, as it makes a very specific prediction for their globular clusters. Both galaxies formed out of the gas that was left behind by the progenitor galaxies at the same time. This gas mixed efficiently during the collision and should have a uniform metallicity. While overall star formation likely lasted for $\gtrsim 500$ Myr (Shin et al. 2020; Lee et al. 2021), the globular clusters formed almost instantaneously before further enrichment could substantially change the abundance of the gas (Lee et al. 2021). As a result, the stellar populations of all the globular clusters, in both galaxies, should be extremely similar.¹⁰ This is not what is typically observed in dwarf galaxies. In a comprehensive study of the colors of globular clusters in Virgo dwarf galaxies, Peng et al. (2006) found that there is substantial cluster-to-cluster and galaxy-to-galaxy scatter.

Previous measurements of the colors of the globular clusters in DF2 and DF4 do in fact suggest differences between them, in potential conflict with the bullet model. Tables 1 and 2 in Shen et al. (2021a) imply a mean $V_{606} - I_{814}$ globular cluster color of 0.37 ± 0.01 AB mag in DF2 and 0.41 ± 0.01 in DF4. This difference seems small, but it corresponds to ≈ 0.3 dex in age or ≈ 0.4 dex in metallicity. Furthermore, the cluster-to-cluster scatter is nonzero, with $\sigma = 0.04 \pm 0.01$ mag for both galaxies. There is also a hint that the galaxies themselves might have different $V_{606} - I_{814}$ colors: Cohen et al. (2018) find 0.40 for the diffuse light in DF2 and 0.32 for DF4, albeit with an uncertainty of 0.1 for both.

These measurements were performed using standard techniques (such as Source Extractor; Bertin & Arnouts 1996) and are largely based on single-orbit HST/ACS images that were reduced with the default STScI flat fields. Here we remeasure the colors of the clusters and the diffuse light in both DF2 and DF4 using custom techniques and well-calibrated, much deeper data, as a stringent test of the bullet dwarf model. Where needed, we use a distance of $D = 21.7$ Mpc for DF2 and $D = 20.0$ Mpc for DF4 (Danieli et al. 2020; Shen et al. 2021b; Z. Shen et al., in preparation).

2. Data

We make use of deep HST/ACS data that were obtained in programs GO-15695 and GO-15851. The aim of these programs was to measure distances to DF4 (GO-15695) and DF2 (GO-15851) from the tip of the red giant branch (TRGB). For DF4, the exposure times were seven and three orbits in I_{814} and V_{606} , respectively. In Danieli et al. (2020), these were combined with the 1+1 orbits that had been obtained in GO-14644 (see Cohen et al. 2018), and the location of the TRGB was measured from these 8+4 orbit depth data. The DF2 observations of GO-15851 were deeper, at 19 orbits in I_{814} and 19 orbits in V_{606} . Shen et al. (2021b) used these data to measure the TRGB distance of DF2, again after adding the GO-14644 data for a total exposure of 20+20 orbits.

In this study we redrizzle the ACS data, with several changes. First, we discard the 1+1 orbits that were obtained for

DF2 and DF4 in GO-14644. The orientation of these observations was very different from the more recent data, which means the point-spread function (PSF) and charge transfer efficiency corrections are different, and there is no need to maximize depth: The fluxes of the globular clusters are typically $\sim 30 e^- s^{-1}$, which means that Poisson errors are $\lesssim 0.5\%$ even in a single orbit.

More importantly, we apply a flat-fielding correction to the `flc` files prior to drizzling. The ACS flat fields were last updated in 2006, and they no longer provide optimal corrections. As described in the instrument science report¹¹ ACS 2017-09, ultradeep stacks of Frontier Fields images show that there are systematic flat-fielding residuals at the level of 1%. This effect is also described in ISR ACS 2020-08. The residuals correlate with the local thickness of the CCD, and as the correlation between quantum efficiency and thickness reverses at ≈ 700 nm, the residuals in V_{606} and I_{814} are spatially anticorrelated. As a result, flat-fielding errors in the $V_{606} - I_{814}$ color reach 2%, even though they are only half that in each filter individually.

The ACS team at STScI provided us with preliminary correction flats, and we applied these to the `flc` files. The `flc` files were aligned with each other using `tweakreg`. The `tweakreg` algorithm is sensitive to cosmic rays; we addressed this by running the code on versions of the data where cosmic rays were removed (with `L.A.Cosmic` van Dokkum 2001). I_{814} images of DF2 and DF4 are shown in Figure 1. Visually comparing the default and flat-field-corrected images, there is a clear improvement in the flatness of the sky background, particularly for DF2. From the remaining variation in the background, we estimate that residual flat-fielding errors are $0.5\% \pm 0.2\%$. Assuming that the errors are no longer (anti) correlated, the uncertainty in the color due to spatially varying flat-fielding uncertainties is then $0.7\% \pm 0.3\%$ in our images.

3. Measurements of Colors and Magnitudes

3.1. Sample

At present, 24 globular clusters have been spectroscopically confirmed, 12 in DF2 and 12 in DF4. These are listed in Tables 1 and 2 of Shen et al. (2021a); the sources of the spectra are van Dokkum et al. (2018a, 2019), Emsellem et al. (2019), and Shen et al. (2021a). The sample is likely complete for $M_V \lesssim -8.5$. Several clusters are not included in our analysis. DF2-80 is superposed on a faint spectroscopically identified blue background galaxy (Shen et al. 2021a) with no possibility of disentangling the two objects; DF4-97 falls outside of the coverage of the GO-15695 data, and DF4-926 is projected onto a bright background merger (see van Dokkum et al. 2019). This leaves 11 confirmed clusters in DF2 and 10 in DF4.

Three of the confirmed DF4 clusters are fainter than $M_V = -8$. As we show later the errors increase sharply at fainter magnitudes, and our quantitative analysis of the $V_{606} - I_{814}$ scatter focuses on the $11 + 7 = 18$ objects that have $M_V < -8.6$. The faint clusters are used to investigate trends with magnitude; to this end, we use an augmented sample in Section 4 that includes five faint objects without a spectrum. These candidate clusters satisfy the Shen et al. (2021a) color and size criteria and are located within $45''$ of the center of DF2 or DF4. The sample is listed in Table 1.

¹⁰ We note here that simulations have so far only focused on comparing the properties of clusters across a single resulting galaxy fragment and not yet across several galaxies.

¹¹ The repository of these reports is <https://www.stsci.edu/hst/instrumentation/acs/documentation/instrument-science-reports-isrs>.

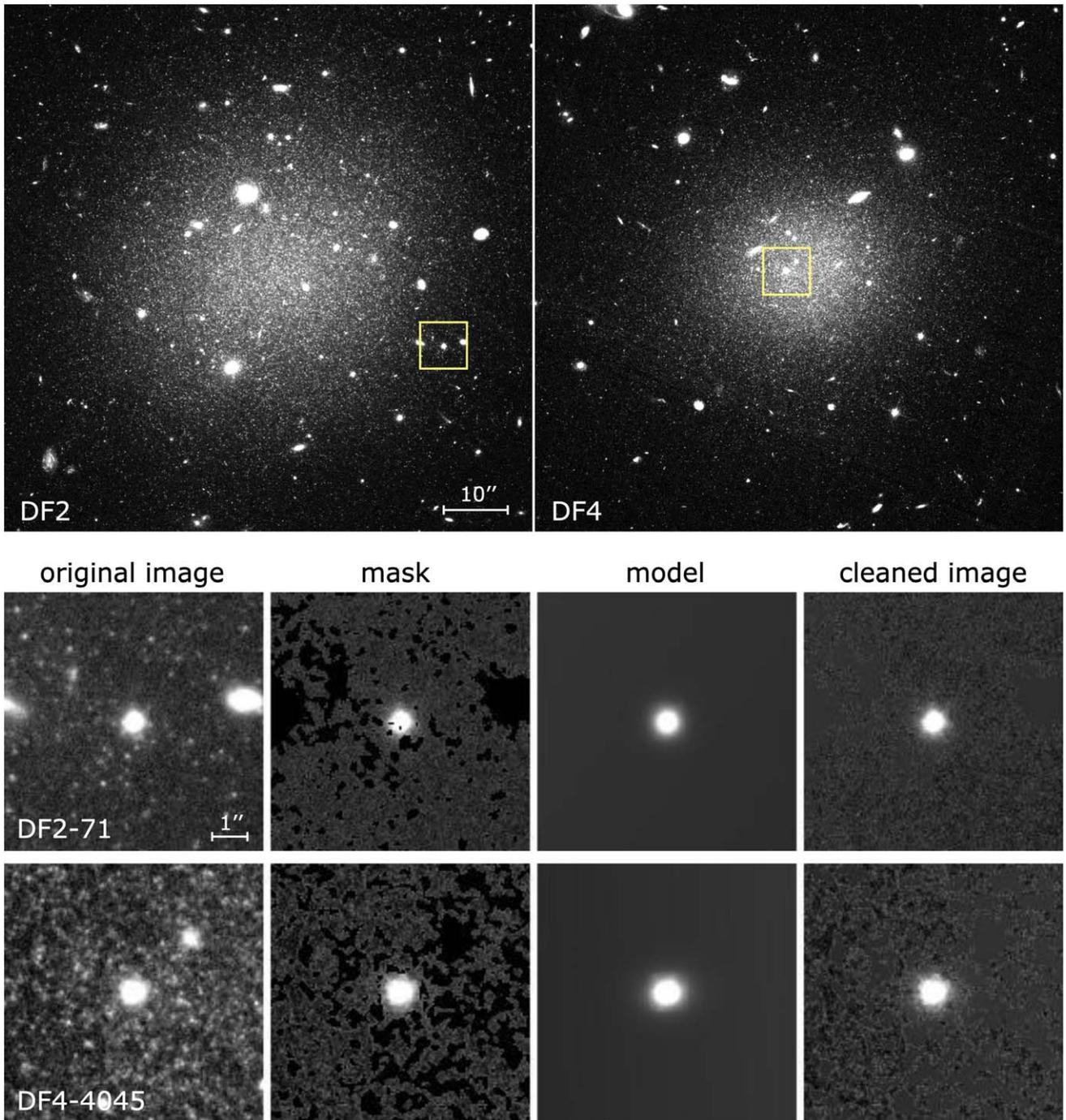


Figure 1. Top panels: images of DF2 and DF4 in the HST/ACS I_{814} filter, after applying flat-field corrections to the `flc` files and drizzling. The images span $80'' \times 80''$. Bottom panels: photometry procedure for two of the globular clusters, indicated with yellow boxes above. From left to right, we show the original I_{814} image (with a size of $7'' \times 7''$), the final object mask, the PSF-convolved King (1962) model fit with the best-fitting background plane, and the cleaned image. The cleaned image is the original image with the masked areas filled in by the King model. Colors and total magnitudes are measured from aperture photometry on the cleaned V_{606} and I_{814} images.

3.2. Cleaned Globular Cluster Images

Even with flat-fielding errors largely eliminated, it is not straightforward to obtain photometry with the desired accuracy. An investigation into the effects of changes in aperture sizes and background subtraction shows seemingly random variation in the colors at the 1%–2% level for individual clusters. These variations can be traced to the cumulative effects of contaminating objects, mostly red giants in the main bodies

of the galaxies, on both the flux measurements and the estimation of the background.

In shallow data, this is an unavoidable source of noise, but owing to the depth of the I_{814} images, we can mitigate these effects. We use the following procedure for each cluster. We begin by creating a mask for the brightest contaminating objects in a $7'' \times 7''$ region centered on the cluster. A simple lower limit of $0.03 \text{ e}^- \text{ s}^{-1} \text{ pix}^{-1}$ is used, excluding the central $1''.0 \times 1''.0$ to avoid masking the cluster itself. Next, an

Table 1
Photometry of Globular Clusters

ID	M_{606}	$V_{606} - I_{814}$	\pm Meas	\pm Phot	\pm Stoch
Main Sample					
DF2-39	-9.40	0.363	0.001	0.007	0.008
DF2-59	-8.97	0.389	0.004	0.008	0.010
DF2-71	-9.12	0.396	0.007	0.010	0.009
DF2-73	-10.24	0.365	0.002	0.007	0.007
DF2-77	-9.76	0.356	0.006	0.009	0.007
DF2-85	-9.35	0.388	0.010	0.012	0.008
DF2-91	-9.29	0.377	0.005	0.008	0.009
DF2-92	-9.61	0.382	0.005	0.009	0.007
DF2-93	-8.89	0.349	0.011	0.013	0.010
DF2-98	-8.97	0.374	0.003	0.008	0.010
DF2-101	-8.88	0.365	0.003	0.008	0.010
DF4-1092	-9.12	0.377	0.002	0.007	0.008
DF4-1403	-8.96	0.384	0.004	0.008	0.009
DF4-3515	-8.67	0.374	0.005	0.009	0.010
DF4-4045	-9.69	0.375	0.006	0.009	0.007
DF4-5675	-9.29	0.360	0.004	0.008	0.008
DF4-6129	-9.13	0.411	0.004	0.008	0.008
DF4-6571	-8.93	0.381	0.004	0.008	0.009
Faint Sample					
DF2-2764	-8.02	0.417	0.026	0.027	0.016
DF2-3555	-7.70	0.391	0.017	0.018	0.017
DF2-4061	-8.02	0.378	0.020	0.021	0.016
DF4-1755	-7.92	0.429	0.015	0.017	0.016
DF4-3278	-8.17	0.324	0.026	0.027	0.013
DF4-3602	-8.57	0.427	0.021	0.022	0.011
DF4-4176	-7.25	0.379	0.070	0.071	0.023
DF4-4366	-7.40	0.429	0.018	0.019	0.020

Note. Magnitudes are on the AB system. IDs are from Shen et al. (2021a). Uncertainties on the colors are $\pm 1\sigma$, with \pm meas the measurement uncertainty, \pm phot the measurement uncertainty combined with the flat-fielding uncertainty, and \pm stoch the stochastic uncertainty due to variations in the number of red giants.

approximate model for the cluster is generated by fitting a PSF-convolved 2D modified King profile with $\alpha = 2$ (Elson 1999; Peng et al. 2002) to the image, using `galfit` (Peng et al. 2002). The bright object mask is used, the background is modeled as a constant, and for the PSF, we choose a non-saturated star in the image.

With a model for the cluster in hand an improved mask can be created by identifying pixels that deviate significantly from the model. Specifically, we create a residual image $R = (I - M)/M_{0.02}$, with I the image, M the model, and $M_{0.02}$ the model with all pixels $\leq 0.02 \text{ e}^- \text{ s}^{-1} \text{ pix}^{-1}$ set to 0.02. After median-filtering R by a 3×3 pixel filter, all pixels > 0.15 are flagged and added to the initial mask. This process masks objects whose median flux in a 3×3 pixel aperture deviates by more than 15% from the model, as well as objects away from the central regions whose flux exceeds a fixed threshold of $\gtrsim 0.003 \text{ e}^- \text{ s}^{-1}$ (with the precise threshold depending on the background level). The chosen values are a compromise between masking as many contaminants as possible and leaving sufficient data for reliable flux and background measurements. The masks for two representative clusters, DF2-71 and DF4-4045, are shown in Figure 1.

The fit to the globular cluster is repeated with this new mask, now modeling the background as a plane with gradients in x

and y to properly account for the diffuse light of DF2 and DF4. The models for DF2-71 and DF4-4045 are shown in the third column of Figure 1. Finally, a cleaned image is created by replacing all masked pixels in the original image by the corresponding pixels in the model (fourth column of Figure 1).

3.3. Aperture Photometry

Fluxes are measured directly from the cleaned images using simple aperture photometry. Colors are measured in $0''.5$ diameter apertures and the total V_{606} magnitudes in $1''.5$ diameter apertures. The background is determined from an annulus with an inner diameter of $1''.5$ and an outer diameter of $3''.0$. As noted above, the results depend on the precise choice of these parameters when applied to the original images, but we find that they are insensitive to them when applied to the cleaned images.

Errors on the fluxes were determined empirically by repeating the aperture photometry in a grid of 6×6 blank positions within the $7'' \times 7''$ cleaned images. This procedure ensures that the local environment of each globular cluster is taken into account. As can be seen in the mask images of Figure 1, the central region (where the globular cluster is) is typically not masked as faint stars do not exceed the thresholds there. This leads to a bias that can be quantified and corrected for in the grid photometry. At each grid position, we replaced the pixels within $\pm 0''.25$ of that position with those from the original image. The grid photometry now closely resembles the globular cluster photometry. The mean flux from the 36 grid positions is subtracted from the globular cluster measurements, and the scatter among them is taken as the uncertainty. Both values were determined with the biweight estimator (Beers et al. 1990) as it is insensitive to outliers.

Aperture corrections are taken from Bohlin (2016); as the half-light radii of the globular clusters are typically $0''.05$ (van Dokkum et al. 2018a), these point-source corrections are adequate for obtaining accurate colors. Fluxes are converted to AB magnitudes using the zero-points in the headers of the images and corrected for Galactic extinction using the Schlafly & Finkbeiner (2011) maps. The corrections are -0.060 mag, -0.062 mag in V_{606} for DF2, DF4, and -0.037 mag, -0.038 mag in I_{814} . Finally, total magnitudes are converted to absolute magnitudes.

4. Observed Color Variation

The distribution of the DF2 and DF4 globular clusters in the color-magnitude plane is shown in the top panel of Figure 2. Error bars are a combination of the measurement uncertainty and the 0.007 mag flat-fielding uncertainty. For reference, the grayscale background shows the parameterized distribution of globular clusters in low-luminosity galaxies in the Virgo cluster. This is a combination of the two-component decomposition of the color distribution for the faintest galaxies¹² in Peng et al. (2006) and the Gaussian fit to the globular cluster luminosity function of the faintest galaxies¹³ in Jordán et al. (2007).

¹² Specifically, we averaged the fit values for the $M_B = -16.6$ and $M_B = -15.7$ bins in Table 4 of Peng et al. (2006). The $g - z$ colors were converted using $V_{606} - I_{814} = 0.445(g - z) + 0.061$.

¹³ We averaged the g -band fit values for the $M_B = -16.6$, $M_B = -16.4$, and $M_B = -15.7$ bins in Table 3 of Jordán et al. (2007) and converted to V_{606} using $V_{606} = g - 0.40$.

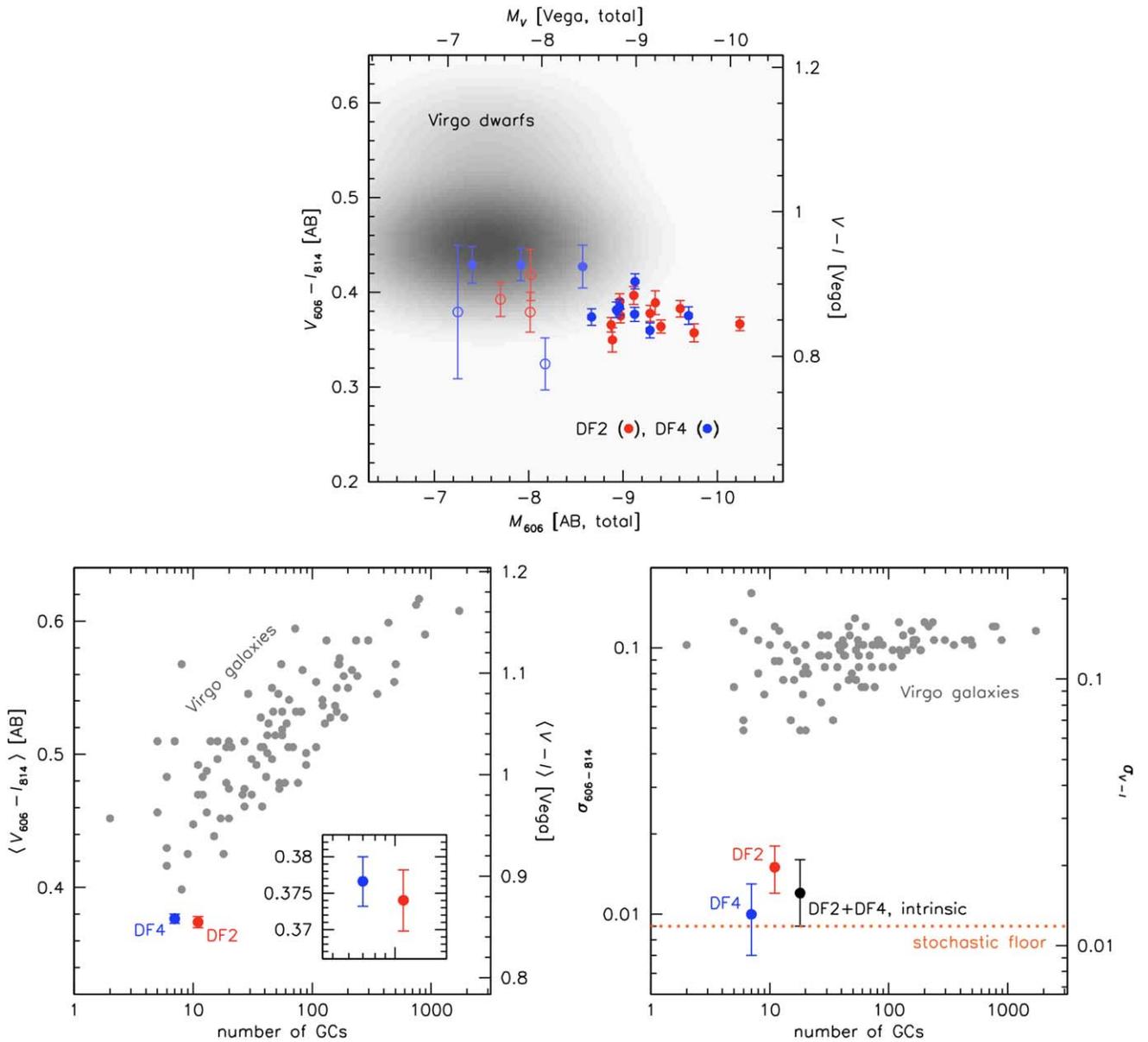


Figure 2. Top panel: DF2/DF4 globular clusters in the color–magnitude plane. Solid symbols are spectroscopically confirmed, and light-colored symbols have uncertainties < 0.015 mag. The parameterized distribution for globular clusters in low-luminosity Virgo galaxies is shown in gray (see text). Bottom-left panel: mean color of globular clusters with uncertainties < 0.015 , compared to Virgo galaxies from Peng et al. (2006). There is no significant systematic color difference between DF2 and DF4. Bottom-right panel: observed and intrinsic scatter in the colors, again compared to Virgo galaxies. The clusters in DF2 and DF4 are extremely homogeneous. The orange line is the minimum possible scatter, arising from stochastic fluctuations in the number of red giants.

The globular clusters in DF2 and DF4 are much brighter than those in Virgo dwarfs, as has been discussed extensively in earlier papers (see van Dokkum et al. 2018a; Trujillo et al. 2019; Shen et al. 2021a), and they are also somewhat bluer. There is no evidence for a systematic trend with magnitude when the full sample of confirmed and candidate clusters is considered. The photometric uncertainties increase sharply at fainter magnitudes, and in the following, we only consider the 18 clusters with errors < 0.015 . This sample corresponds to the full sample of clusters with $M_{606} < -8.6$.

The mean color of the DF2 and DF4 clusters is nearly identical: $\langle V_{606} - I_{814} \rangle = 0.374 \pm 0.004$ for DF2 and

$\langle V_{606} - I_{814} \rangle = 0.377 \pm 0.003$ for DF4,¹⁴ as determined with the biweight estimator (Beers et al. 1990). The mean color difference is $\Delta_{DF2-DF4} = -0.003 \pm 0.005$. The mean colors are compared to those of globular clusters in Virgo galaxies¹⁵ in Figure 2 (lower left). The clusters in DF2 and DF4 are bluer than those in Virgo galaxies with similar N_{GC} .¹⁶ Furthermore, in the Virgo sample with $N_{GC} < 20$ the median color difference

¹⁴ The error bars do not include an uncertainty of $\approx 1\%$ in the absolute calibration of the ACS filters, as this systematic error affects all colors and magnitudes by the same amount.

¹⁵ The Virgo data are the single-component fits in Table 3 of Peng et al. (2006).

¹⁶ The color difference partly reflects an age difference: The DF2 clusters have ages of 9 ± 2 Gyr (van Dokkum et al. 2018a; Fensch et al. 2019) whereas the typical ages of metal-poor globular clusters are similar to those in the Milky Way (see Strader et al. 2005).

between any two data points is 0.036, an order of magnitude larger than the difference between DF2 and DF4.

The cluster-to-cluster scatter is also very small and within the errors is the same for the two galaxies: $\sigma_{\text{obs}} = 0.015 \pm 0.003$ for DF2 and $\sigma_{\text{obs}} = 0.010 \pm 0.003$ for DF4. The observed scatter in the combined sample of 18 bright globular clusters in DF2 and DF4 is $\sigma_{\text{obs}} = 0.015 \pm 0.002$ (all determined with the biweight estimator; the rms is also 0.015). The intrinsic scatter σ_{intr} can be determined by constructing the likelihood function,

$$\mathcal{L} = \prod_{i=1}^{i \leq 18} \frac{1}{\sqrt{2\pi} \sigma_{\text{eff}}} \exp \left[-0.5 \left(\frac{c_i - \mu}{\sigma_{\text{eff}}} \right)^2 \right], \quad (1)$$

with c_i the colors of the individual clusters, μ the mean, and $\sigma_{\text{eff}}^2 = \sigma_{\text{intr}}^2 + e_i^2$. We find an intrinsic scatter of $\sigma_{\text{intr}} = 0.012_{-0.003}^{+0.004}$. As shown in Figure 2 (lower right), the typical scatter in Virgo galaxies is $\sigma \approx 0.1$ mag, and there are no galaxies with $\sigma < 0.05$ mag. We conclude that the globular clusters in DF2 and DF4 form a remarkably homogeneous population, as predicted by the bullet dwarf collision model. This level of homogeneity is not observed in normal dwarf galaxies.

5. Interpretation of the Variation

The color variation is small but not zero. As noted above, observational uncertainties explain some of the observed scatter. This is shown in the top-left panel of Figure 3, where the error bar for each data point is split into several distinct contributions. Measurement uncertainties are shown in black and the total photometric uncertainties, which include the 0.007 ± 0.003 mag flat-fielding errors, are shown in dark gray.

The intrinsic scatter is so small that the effects of stochastic sampling of the isochrone need to be taken into account. Red giants contribute significantly to the integrated light, and the Poisson variation in their number causes some clusters to be redder and brighter than others. The same effect causes the well-known pixel-to-pixel surface brightness fluctuations in galaxy images (Tonry & Schneider 1988; Greco et al. 2021). We quantify this effect by generating artificial globular clusters with the `ArtPop` code (Greco & Danieli 2021) and measuring their integrated colors. The clusters have $\log(\text{age}) = 9.9$, $-1.8 \leq [\text{Fe}/\text{H}] \leq -0.8$, and cover a factor of 20 in mass (parameterized by $\log(n_{\text{stars}})$, which ranges from 5 to 6.3 with steps of 0.2). At each mass, 500 clusters are generated with different random seeds, and the rms scatter in the $V_{606} - I_{814}$ colors is measured.

The results are shown in the bottom-right panel of Figure 3. The cluster-to-cluster $V_{606} - I_{814}$ scatter is $\sim 1\%$ in the relevant luminosity range, of the same order as the intrinsic scatter in the DF2/DF4 globular clusters, with a modest dependence on metallicity. Two example clusters that illustrate the effect are shown at left (see also Figure 6 in Greco & Danieli 2021). A polynomial fit to the luminosity-dependent variation has the form

$$\sigma_{\text{stoch}} = 0.019 + 0.0087(M_{606} + 7.5) + 0.00153(M_{606} + 7.5)^2 \quad (2)$$

and is shown by the black line in Figure 3.

Light-gray error bars in the top panel of Figure 3 show the effect of including this uncertainty for each cluster. The broken orange lines show where 68% of the points are expected to fall

due to the stochastic sampling effect alone. Solid orange lines include the measurement error; for a normal distribution entirely defined by these lines, 12 out of 18 would fall within them. They encompass 10, only slightly fewer. Using the total errors (that is, the combination of the measurement error, the flat-fielding error, and the stochastic variation) in the likelihood analysis gives the “stellar population scatter”, $\sigma_{\text{sp}} = 0.008_{-0.006}^{+0.005}$. This scatter is not significantly different from zero. The likelihood function, marginalized over μ , is shown by the solid line in Figure 3.

We use the flexible stellar population synthesis (FSPS) framework (Conroy et al. 2009) with the MIST isochrones (Choi et al. 2016) to determine the (limits on) variation in age and metallicity that is implied by the stellar population scatter. For ages near 8 Gyr and metallicities near $[\text{Fe}/\text{H}] = -1.3$, we find $\Delta(V_{606} - I_{814}) = 0.10 \Delta[\text{Fe}/\text{H}]$ and $\Delta(V_{606} - I_{814}) = 0.0064 \Delta(\text{age})$ to good approximation, with age in Gyr. The observed stellar population scatter σ_{sp} implies $\sigma_{[\text{Fe}/\text{H}]} = 0.08_{-0.06}^{+0.05}$ if there is no variation in age or $\sigma_{\text{age}} = 1.3_{-1.0}^{+0.8}$ Gyr if there is no variation in metallicity. In the simulations of Lee et al. (2021), the globular clusters can have a spread of up to ≈ 0.1 dex in metallicity and ≈ 150 Myr in age, and we conclude that the bullet hypothesis cannot be ruled out.

6. Colors of the Diffuse Light

The bullet model also makes predictions for the global colors of DF2 and DF4, although these are more model dependent than the predictions for the globular clusters (Shin et al. 2020; Lee et al. 2021). The galaxies are predicted to have higher metallicities than the globular clusters as they form stars over a longer time period (see Lee et al. 2021). Therefore, while the colors of DF2 and DF4 should be very similar to each other, they are predicted to be redder than those of the globular clusters.

We measure the average colors of the galaxies in the following way. An object mask is created by comparing the summed $V_{606} + I_{814}$ image to a binned and median-filtered version of itself. A first model for the galaxy is made by median-filtering the V_{606} and I_{814} images, not taking masked pixels into account. This model is subtracted from the data, and the object mask is optimized with a lower threshold, taking care not to include giants in the mask. Then, the median-filtering is repeated to create a final model in each filter. There is a background gradient in all images, and at this stage, a surface is fitted to the background and subtracted. The rms of this background model (that is, 68% of the gradient that is removed) is taken as the uncertainty in each filter. Finally, the I_{814} model is divided by the V_{606} model, multiplied by the object mask, and a flux threshold is applied so that the faint outskirts are excluded.

These color images are shown in Figure 4 and compared to the mean color of the globular clusters. The variations within each galaxy are caused by stochastic variation in the number of red giants (surface brightness fluctuations). We measure the mean colors of the galaxies to be $(V_{606} - I_{814})_{\text{DF2}} = 0.420 \pm 0.024$ and $(V_{606} - I_{814})_{\text{DF4}} = 0.436 \pm 0.075$, consistent with the Cohen et al. (2018) measurements with smaller uncertainties. We infer that the colors of the galaxies are identical within the errors, as predicted by the bullet model.

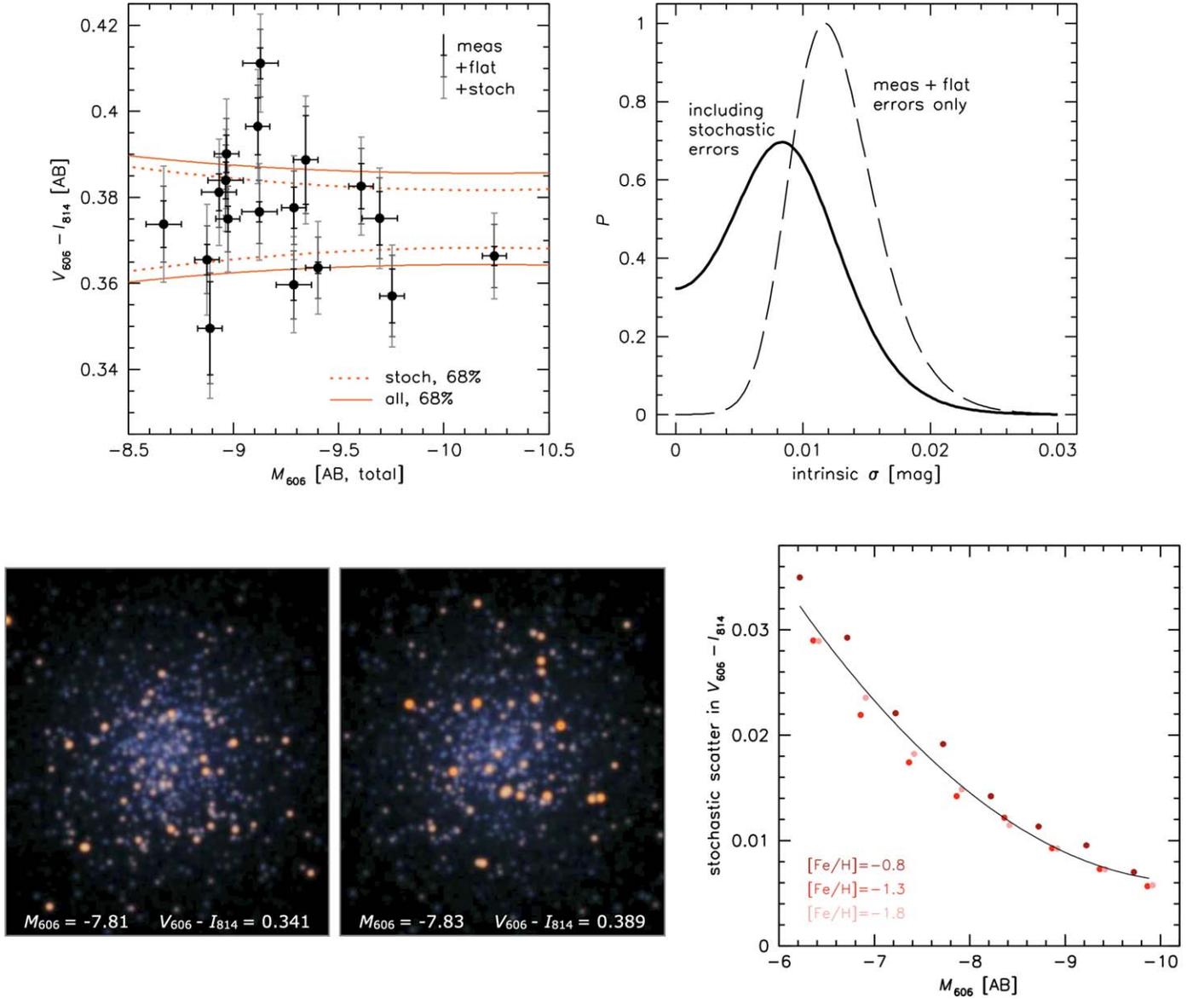


Figure 3. Top-left panel: zoom-in on color–magnitude measurements with errors <0.015 , corresponding to all clusters with $M_{606} < -8.6$. Error bars are the measurement uncertainty (black), the combined measurement and flat-fielding uncertainty (dark gray), and the combined measurement, flat-fielding, and stochastic sampling uncertainty (light gray). Dotted orange lines show the expected $\pm 1\sigma$ variation due to stochastic sampling alone. Solid lines include the other uncertainties. Top-right panel: likelihood for the intrinsic dispersion, for the photometric errors (broken line), and for the total error, including the effect of stochastic sampling (solid line). Bottom: the effects of stochastic sampling of the isochrone. Intrinsic cluster-to-cluster color differences of 0.01–0.02 mag arise from random variations in the number of red giants. Two examples of identical clusters with different random seeds are shown on the left (see the text for details).

The galaxies are redder than the luminous globular clusters. If the actual color of both galaxies is $V_{606} - I_{814} = 0.375$ and the errors are Gaussian, the probability that we measure $(V_{606} - I_{814})_{\text{DF2}} \geq 0.420$ and $(V_{606} - I_{814})_{\text{DF4}} \geq 0.436$ is $<1\%$. The color difference between the galaxies and the clusters of 0.05 mag is qualitatively consistent with the metallicity difference predicted by the hydrodynamical model of Lee et al. (2021). Using the relation in Section 5, we find $\Delta[\text{Fe}/\text{H}] \approx 0.5$, somewhat larger than the ≈ 0.2 predicted by Lee et al. (2021) but in good agreement with the spectroscopically determined value of $\Delta[\text{Fe}/\text{H}] = 0.56 \pm 0.15$ of Fensch et al. (2019).

7. Discussion

The central result of this paper is that the bright globular clusters in DF2 and DF4 have extremely similar colors. The observed scatter is ≈ 0.015 mag, and this can be explained by a combination of measurement uncertainties (≈ 0.01) and stochastic variations in the number of red giants (≈ 0.01). The remaining scatter among the 18 luminous clusters in DF2 and DF4 is $\sigma_{\text{ssp}} = 0.008^{+0.005}_{-0.006}$, that is, not significantly different from zero. The diffuse light is redder in both galaxies, with again no significant difference between DF2 and DF4. These results are expected in the bullet dwarf scenario (see Silk 2019; Shin et al. 2020; Lee et al. 2021; van Dokkum et al. 2022), and we conclude that this model survives an important falsification test. Returning to the formation models listed in Section 1, no other published explanation for the lack of dark matter in DF2

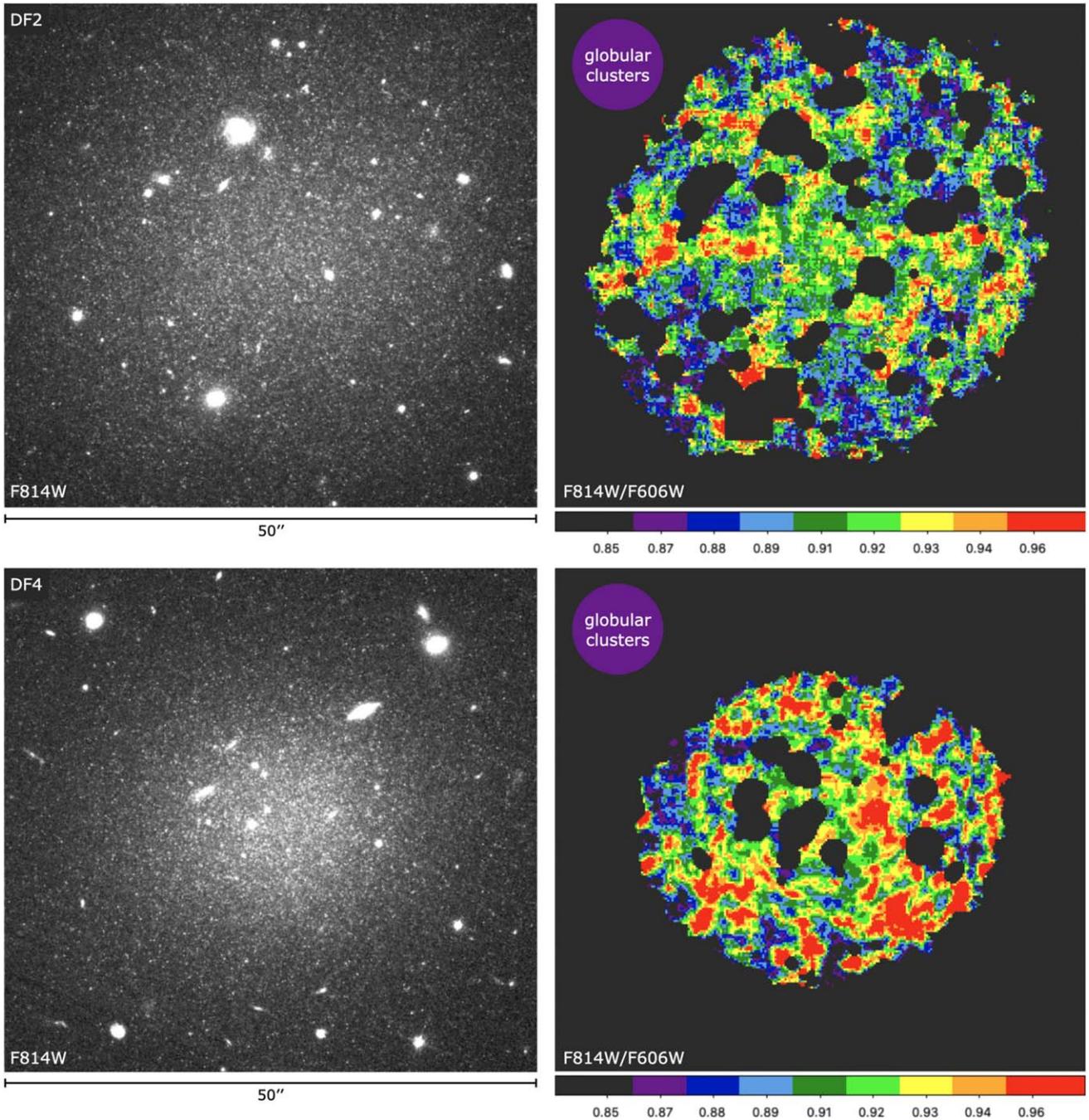


Figure 4. Color images of the galaxies, created by dividing binned and median-filtered I_{814} and V_{606} images of the galaxies. Numerical values are simply the ratios of the observed fluxes in $e^- s^{-1}$. The mean value for the globular clusters is indicated with the circle. Color variations within each galaxy are caused by surface brightness fluctuations. The average color of DF4 is slightly redder than that of DF2 but the difference is not significant: $(V_{606} - I_{814})_{DF2} = 0.420 \pm 0.024$ and $(V_{606} - I_{814})_{DF4} = 0.436 \pm 0.075$. The galaxies are redder than the globular clusters.

and DF4 also “naturally” produces the extreme uniformity of their globular cluster populations.

The globular clusters in DF2 and DF4 are different from those in Virgo galaxies; as detailed in Section 4, they are brighter, bluer, and have a much smaller scatter. DF2 and DF4 are ultradiffuse galaxies (UDGs; van Dokkum et al. 2015) and at least two other UDGs also have very homogeneous globular cluster populations, NGC 5846-UDG1 (Müller et al. 2021; Danieli et al. 2022) and DGSAT1 (Janssens et al. 2022). In other respects they are different; NGC 5846-UDG1 has a canonical globular cluster luminosity function, and DGSAT1’s

clusters are significantly redder than those in DF2/DF4. A small scatter indicates synchronized formation in a dense, homogeneous medium, and it may arise generically in any formation scenario that results in the extreme (factor of 10^9) density contrast between the globular clusters and the galaxy light that is seen in UDGs (see, e.g., van Dokkum et al. 2018a; Trujillo-Gomez et al. 2021). Specifically, intense feedback from the formation of the globular clusters may have caused the galaxies to expand and turn into UDGs (Trujillo-Gomez et al. 2021; Danieli et al. 2022). Further deep HST or JWST studies

of globular clusters in UDGs in various environments are needed to investigate this further.

Our analysis focuses on the brightest clusters as these have the smallest uncertainties. Shen et al. (2021a) showed that the globular cluster luminosity function in DF2 and DF4 can be modeled as a combination of a bright peak of overluminous clusters plus a “normal” luminosity function with a normalization and mean that are typical for the galaxies’ luminosities. In this context, it is interesting that the clusters with $M_{606} > -8.6$ are somewhat redder than the brighter ones, with $\langle V_{606} - I_{814} \rangle = 0.400 \pm 0.011$. Perhaps the faint clusters formed together with the diffuse light, with a higher mean metallicity than the brighter ones.¹⁷ The observed scatter is also higher for the faint clusters at $\sigma_{606-814} = 0.034 \pm 0.008$, although this can be explained by a combination of measurement errors (0.028) and stochastic sampling (0.016). Extremely deep spectroscopy may shed more light on these questions.

Further tests of the bullet model are possible. The most straightforward next step is probably obtaining radial velocities and line-of-sight distances of other galaxies along the trail, as the bullet model predicts that these follow a regular sequence (with some contamination from unrelated objects). Dynamical mass measurements of other trail galaxies will also be highly constraining and interesting in a broader context as bullet dwarf events can, in principle, constrain the self-interaction cross section of dark matter. Modeling of the bullet cluster has provided an upper limit (Randall et al. 2008), but as self-interacting dark matter was introduced to explain the “cored” dark matter density profiles of low-mass galaxies (Spergel & Steinhardt 2000), it is important to measure the cross section on those scales (Tulin & Yu 2018).

We thank Yotam Cohen and the rest of the STScI ACS team for their help with the data reduction. Support from HST grants GO-14644, GO-15695, and GO-15851 is gratefully acknowledged. S.D. is supported by NASA through Hubble Fellowship grant HST-HF2-51454.001-A.

ORCID iDs

Pieter van Dokkum  <https://orcid.org/0000-0002-8282-9888>
 Zili Shen  <https://orcid.org/0000-0002-5120-1684>
 Aaron J. Romanowsky  <https://orcid.org/0000-0003-2473-0369>
 Roberto Abraham  <https://orcid.org/0000-0002-4542-921X>
 Charlie Conroy  <https://orcid.org/0000-0002-1590-8551>
 Shany Danieli  <https://orcid.org/0000-0002-1841-2252>
 Dhruva Dutta Chowdhury  <https://orcid.org/0000-0003-0250-3827>
 Michael A. Keim  <https://orcid.org/0000-0002-7743-2501>
 J. M. Diederik Kruijssen  <https://orcid.org/0000-0002-8804-0212>
 Joel Leja  <https://orcid.org/0000-0001-6755-1315>
 Sebastian Trujillo-Gomez  <https://orcid.org/0000-0003-2482-0049>

References

Beers, T. C., Flynn, K., & Gebhardt, K. 1990, *AJ*, 100, 32

- Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
 Bohlin, R. C. 2016, *AJ*, 152, 60
 Carleton, T., Errani, R., Cooper, M., et al. 2019, *MNRAS*, 485, 382
 Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, 823, 102
 Cohen, Y., van Dokkum, P., Danieli, S., et al. 2018, *ApJ*, 868, 96
 Conroy, C., Gunn, J. E., & White, M. 2009, *ApJ*, 699, 486
 Danieli, S., van Dokkum, P., Abraham, R., et al. 2020, *ApJL*, 895, L4
 Danieli, S., van Dokkum, P., Conroy, C., Abraham, R., & Romanowsky, A. J. 2019, *ApJL*, 874, L12
 Danieli, S., van Dokkum, P., Trujillo-Gomez, S., et al. 2022, *ApJL*, 927, L28
 Dutta Chowdhury, D., van den Bosch, F. C., & van Dokkum, P. 2019, *ApJ*, 877, 133
 Elson, R. A. W. 1999, in *Canary Islands Winter School of Astrophysics: Globular clusters*, 10 (Cambridge: Cambridge Univ. Press), 209
 Emsellem, E., van der Burg, R. F. J., Fensch, J., et al. 2019, *A&A*, 625, A76
 Fensch, J., van der Burg, R. F. J., Jeřábková, T., et al. 2019, *A&A*, 625, A77
 Greco, J. P., & Danieli, S. 2021, *ApJ*, submitted, arXiv:2109.13943
 Greco, J. P., van Dokkum, P., Danieli, S., Carlsten, S. G., & Conroy, C. 2021, *ApJ*, 908, 24
 Janssens, S. R., Romanowsky, A. J., Abraham, R., et al. 2022, *MNRAS*, 517, 858
 Jordán, A., McLaughlin, D. E., Côté, P., et al. 2007, *ApJS*, 171, 101
 Keim, M. A., van Dokkum, P., Danieli, S., et al. 2022, *ApJ*, 935, 160
 King, I. 1962, *AJ*, 67, 471
 Laporte, C. F. P., Agnello, A., & Navarro, J. F. 2019, *MNRAS*, 484, 245
 Lee, J., Shin, E.-j., & Kim, J.-h. 2021, *ApJL*, 917, L15
 Ma, J., Wang, S., Wang, S., et al. 2020, *MNRAS*, 496, 3741
 Martin, N. F., Collins, M. L. M., Longeard, N., & Tollerud, E. 2018, *ApJL*, 859, L5
 Montes, M., Infante-Sainz, R., Madrigal-Aguado, A., et al. 2020, *ApJ*, 904, 114
 Moreno, J., Danieli, S., Bullock, J. S., et al. 2022, *NatAs*, 6, 496
 Müller, O., Durrell, P. R., Marleau, F. R., et al. 2021, *ApJ*, 923, 9
 Natarajan, P., Sigurdsson, S., & Silk, J. 1998, *MNRAS*, 298, 577
 Nusser, A. 2020, *ApJ*, 893, 66
 Ogiya, G. 2018, *MNRAS*, 480, L106
 Ogiya, G., van den Bosch, F. C., & Burkert, A. 2022, *MNRAS*, 510, 2724
 Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, *AJ*, 124, 266
 Peng, E. W., Jordán, A., Côté, P., et al. 2006, *ApJ*, 639, 95
 Randall, S. W., Markevitch, M., Clowe, D., Gonzalez, A. H., & Bradač, M. 2008, *ApJ*, 679, 1173
 Schlafly, E. F., & Finkbeiner, D. P. 2011, *ApJ*, 737, 103
 Shen, Z., van Dokkum, P., & Danieli, S. 2021a, *ApJ*, 909, 179
 Shen, Z., Danieli, S., van Dokkum, P., et al. 2021b, *ApJL*, 914, L12
 Shin, E.-j., Jung, M., Kwon, G., et al. 2020, *ApJ*, 899, 25
 Silk, J. 2019, *MNRAS*, 488, L24
 Spergel, D. N., & Steinhardt, P. J. 2000, *PhRvL*, 84, 3760
 Strader, J., Brodie, J. P., Cenarro, A. J., Beasley, M. A., & Forbes, D. A. 2005, *AJ*, 130, 1315
 Tonry, J., & Schneider, D. P. 1988, *AJ*, 96, 807
 Trujillo, I., Beasley, M. A., Borlaff, A., et al. 2019, *MNRAS*, 486, 1192
 Trujillo-Gomez, S., Kruijssen, J. M. D., Keller, B. W., & Reina-Campos, M. 2021, *MNRAS*, 506, 4841
 Trujillo-Gomez, S., Kruijssen, J. M. D., & Reina-Campos, M. 2022, *MNRAS*, 510, 3356
 Tulin, S., & Yu, H.-B. 2018, *PhR*, 730, 1
 van Breugel, W., Filippenko, A. V., Heckman, T., & Miley, G. 1985, *ApJ*, 293, 83
 van Dokkum, P., Danieli, S., Abraham, R., Conroy, C., & Romanowsky, A. J. 2019, *ApJL*, 874, L5
 van Dokkum, P., Cohen, Y., Danieli, S., et al. 2018a, *ApJL*, 856, L30
 van Dokkum, P., Danieli, S., Cohen, Y., et al. 2018b, *Natur*, 555, 629
 van Dokkum, P., Shen, Z., Keim, M. A., et al. 2022, *Natur*, 605, 435
 van Dokkum, P. G. 2001, *PASP*, 113, 1420
 van Dokkum, P. G., Abraham, R., Merritt, A., et al. 2015, *ApJL*, 798, L45
 Wasserman, A., Romanowsky, A. J., Brodie, J., et al. 2018, *ApJL*, 863, L15

¹⁷ Lee et al. (2021) predict the opposite trend, with the brighter clusters having a higher metallicity.